

Two Sided Markets with Substitution: Mobile Termination Revisited*

Jerry Hausman[†] and Julian Wright[‡]

June 2006

Abstract

The existing analysis of mobile termination is revisited to take into account that mobile subscribers can substitute cheaper mobile-to-mobile calls for expensive fixed-to-mobile calls. We show how this substitution, which has been ignored in the existing literature and regulatory proceedings, can undermine the normal argument of a competitive bottleneck in mobile termination. Termination charges, in equilibrium, are constrained by the ability of consumers to substitute. Moreover, it becomes possible for the privately set termination charges to be too low. A calibrated model of the Australian market shows that cost-based regulation lowers welfare compared to the (non-regulated) equilibrium outcome.

1 Introduction

In the emerging literature on two-sided markets (see Armstrong, 2006, Armstrong and Wright, 2006, Caillaud and Jullien, 2001, 2003, Rochet and Tirole 2003, 2006 among others), a central question of interest is how do platforms set the structure of prices across the two sides of the business, and associated with this, whether the resulting structure of prices is inefficiently distorted. This interest is motivated by the observation that in many such businesses, one side tends to pay disproportionately more than the other, and that in some cases regulators have stepped in to try to re-balance the structure of prices.

*We are very grateful to Jose Miguel Abito for excellent research assistance.

[†]MIT: jhausman@mit.edu

[‡]National University of Singapore: jwright@nus.edu.sg

One important industry where these issues directly arise is in mobile telephony. Concerns about the high level of the charges for terminating fixed-to-mobile calls in those countries where such charges were not regulated, and the resulting high fixed-to-mobile retail prices, have led to widespread regulation of mobile termination charges. Examples of affected countries include Australia, France, Germany, Italy, Japan, United Kingdom among many other countries.¹ However, above-cost mobile termination charges also imply lower charges for mobile subscribers as mobile operators have a greater incentive to attract mobile subscribers to earn this subsidy (the so-called waterbed effect).² Thus, the two-sided market literature's question of the equilibrium versus efficient price structure naturally arises for mobile termination.

Previous theoretical works to address this question include Armstrong (1997, 2002), Gans and King (2000), Thompson *et al.* (2006), Valletti and Houpis (2005) and Wright (1999, 2002). This body of work broadly argues (1) even highly competitive mobile operators will set termination charges to maximize monopoly termination profit and thereby the subsidy they can offer to their subscribers; (2) this competitive bottleneck means equilibrium termination charges are always set too high in equilibrium; and (3) yet the socially optimal termination charge can still be above cost as additional mobile subscribers provide positive externalities to fixed-line callers.

All of these works treat the two groups — fixed line callers on the one hand and mobile subscribers on the other — as two distinct groups. In this paper we move the literature closer to reality by taking into account that fixed-line callers may also themselves be mobile subscribers, and that such subscribers can make mobile-to-mobile calls in addition to and instead of fixed-to-mobile calls. The main conclusions we reach are then (1) mobile operators will set termination charge below the monopoly level; (2) equilibrium termination charges are not *necessarily* too high; and (3) some new network externalities arise that mean

¹In some countries like the United States, where mobile receivers pay for calls, termination charges are indirectly regulated to very low levels (see Littlechild, 2006 for a detailed explanation).

²This effect has been observed in the U.K. after the regulatory imposition of reduced termination rates in July of 2003. Over the period Q1 2003 to Q2 2004, U.K. mobile subscription prices increased by approximately 6.4%. This compares to a decrease in the BLS CPI for cellular telephone in the U.S. of approximately 2%. See Ofcom website, "The Communications Market," Appendices for August and October 2004 and US Bureau of Labor Statistics (BLS) CPI for wireless telephone services (CUUR0000SEED03).

socially optimal charges can be further above cost than otherwise would be the case.

The reasons for these new results stem largely from the substitution possibilities that arise once we account for the fact fixed-line callers may also be mobile subscribers and make mobile-to-mobile calls, the most obvious of which is that if fixed-to-mobile calls are expensive, people that subscribe to a mobile network and wish to call someone on-the-go from home may do so using their mobile phone.³ This consideration is likely to be important since mobile penetration levels typically exceed 70% in most OECD countries, so most callers can indeed substitute in this way if they wish. Moreover, typically average (and certainly on-net) mobile-to-mobile prices are substantially lower than fixed-to-mobile prices in the absence of regulation.⁴ As a result, competing mobile operators will tend to limit their termination charges to avoid too many subscribers substituting to cheaper mobile-to-mobile calls given such calls do not generate termination profits.

The ability to use mobile phones to reach people on-the-go in turn affects the demand for mobile subscription since when fixed-to-mobile prices are high, more people will want to get a mobile phone to avoid these high prices. This implies that setting termination charges above the level which maximizes the subsidy to subscribers (the equilibrium level), will continue to increase mobile penetration. Since mobile subscribers generate positive externalities to other callers (both those calling from home as well as those calling from on-the-go using their mobile phones), this can create additional positive welfare effects of setting above-cost mobile termination charges. The socially optimal charge can be above that set by operators in equilibrium.

Given the growing amount of regulatory intervention in setting mobile termination fees, the economic analysis in this paper should be useful in determining the economic factors that affect termination fees in competitive markets. In this regard, we calibrate our model to data from Australia to obtain precise implications. The equilibrium level of the termination charge predicted by our model is somewhat above the 21 Australian cents observed in June 2004, consistent with the idea that termination charges were constrained by the threat of regulation. However, it is well below the monopoly (or competitive bot-

³Another form of substitution in practice is to text-messaging. Our model would equally apply if text-messaging is considered instead of, or in addition to, mobile-to-mobile calls.

⁴For example, in 2004 the average price of fixed-to-mobile calls was 33 cents per minute in Australia and 42 cents per minute in New Zealand. The corresponding average mobile-to-mobile prices were 10 cents and 16 cents.

tleneck) level that is predicted by existing models given our calibrated parameter values. Although the welfare maximizing level is found to be lower than the equilibrium level, it is still more than three times cost. Notably, regulating termination charges at cost lowers welfare compared to the equilibrium level.

The rest of the paper proceeds as follows. Section 2 lays out our new framework. In Section 3 we derive general results while Section 4 presents results based on a calibration of the model to Australian data. Finally, Section 5 concludes.

2 The model and preliminaries

To endogenize the choice of technology (fixed or mobile) for any given call, the following simple framework is adopted. Suppose that a fraction λ of the time people are at “home” and a fraction $1 - \lambda$ of the time people are “on-the-go”. People at home are assumed to have access to the fixed-line network while when they are “on-the-go” they do not. This gives rise to several potential types of calls — λ^2 potential calls between people at home, $(1 - \lambda)^2$ potential calls between people on-the-go, and $\lambda(1 - \lambda)$ potential calls in each direction between people at home and on-the-go. Whether these calls can be realized or not will depend in part on whether a caller is a mobile subscriber.

In two of these situations, calls are to people at home. To reduce the number of different types of calls under consideration, we assume all such calls are received by people on their landlines and so do not involve mobile termination. This is consistent with the prices of calls to fixed networks being cheaper. Since they just shift consumer utility in an additive way, they are ignored. Focusing on calls to people on-the-go, either these originate from people at home (λ of the time) or originate from people who are also on-the-go ($1 - \lambda$ of the time). In either case, calls are only realized when the receiver is a mobile subscriber. When the caller is also on-the-go, the call is only realized when both parties are mobile subscribers. When instead the caller is at home, the caller may use her landline, or if she is a mobile subscriber, may choose between using her landline or her mobile phone to make the call. This decision is endogenized in the model.

Denoting the per-minute price of fixed-to-mobile (*FTM*) calls p_f and the per-minute price of mobile-to-mobile (*MTM*) calls p_m , one can denote the indirect utility associated with each call from home to

a person on-the-go as $v(p_f, p_m)$ for mobile subscribers who have a choice of which technology to use and $v_f(p_f)$ for non-subscribers who can only make *FTM* calls. The indirect utility associated with each call from someone on-the-go to another person on-the-go is denoted $v_m(p_m)$ since such a call can only be a *MTM* call. Associated with these indirect utility functions are the demand functions $\bar{q}_f(p_f, p_m)$ and $\bar{q}_m(p_f, p_m)$ where callers can choose between *FTM* and *MTM* calls, and $q_f(p_f)$ and $q_m(p_m)$ where callers *cannot* choose the particular technology to use. The demand functions measure the number of minutes people want to spend talking to each mobile subscriber, for each type of call. Taking into account the proportion of time a caller finds himself either at home or on-the-go, we can define the total number of minutes called to each mobile subscriber as $\bar{q}_f = \lambda \bar{q}_f(p_f, p_m)$, $\bar{q}_m = \lambda \bar{q}_m(p_f, p_m)$, $q_f = \lambda q_f(p_f)$ and $q_m = (1 - \lambda) q_m(p_m)$, and the corresponding indirect utility as $v = \lambda v(p_f, p_m)$, $v_f = \lambda v_f(p_f)$ and $v_m = (1 - \lambda) v_m(p_m)$.

Some structure can be imposed on these functions. It must be that $v \geq v_f$ since having the choice of which technology to use cannot make callers worse off, with $v(p_f, \infty) = v_f(p_f)$. In the extreme case in which both technologies are perfect substitutes, then $v(p_f, p_m) = \max(v_f(p_f), v_m(p_m))$ since people will just use the technology which is cheaper. More generally, people will sometimes make use of each technology, reflecting that in different circumstances one technology may be preferred over another.

In particular, assume that over the relevant range of prices, there is positive demand for each type of call, that the own-price derivatives are negative in each case, that the cross price derivatives are positive (i.e. $\partial \bar{q}_f / \partial p_m > 0$ and $\partial \bar{q}_m / \partial p_f > 0$), that $q_f > \bar{q}_f$ so there will be fewer *FTM* calls when a caller has the option of also using their mobile phone to make the calls, and that

$$\left| \frac{\partial \bar{q}_f}{\partial p_f} \frac{p_f}{\bar{q}_f} \right| > \left| \frac{\partial q_f}{\partial p_f} \frac{p_f}{q_f} \right|, \quad (1)$$

so that the price elasticity of *FTM* calls with respect to p_f is higher when people have a choice of which technology to use than when they do not. A given percentage increase in p_f will cause a greater percentage reduction in *FTM* demand by mobile subscribers given they will now partially switch to using their mobile phone for such calls, in addition to making fewer (or shorter) calls in the case they do not switch. All these properties are met, for instance, by a generalized CES specification. They are also met by a linear specification, as illustrated in Section 4, provided parameter values are such that all demands

remain positive.⁵

People also get some other (exogenous) benefits from subscribing to a mobile operator. These benefits include, among other factors, the benefits people get from being able to call people at home when on-the-go, the other services commonly available with mobile subscription, the option value of being able to be reached if necessary or to make an emergency call, and even perhaps the value of receiving calls (if subscribers view this in a lump-sum manner). These benefits are allowed to vary across the population, so that some people are more likely to subscribe than others.⁶ The benefits are denoted b , which are drawn independently for each person from the distribution function G , with corresponding density function g and hazard function $\rho = g/(1 - G)$. These functions are assumed to be continuously differentiable with $g > 0$ over $[\underline{b}, \bar{b}]$, and equal to 0 elsewhere. The total population is of measure one.

If the total measure of people that subscribe to one of the mobile operators is denoted N , a person's utility can be summarized as follows. If a person does not have a mobile subscription, their utility arising from conversing with people on-the-go is just

$$Nv_f. \tag{2}$$

In this case, people only get utility from *FTM* calls, this utility being proportional to the measure of consumers who have subscribed to one of the mobile operators. If instead, a person subscribes to a mobile operator i , their utility becomes

$$b - r_i + N(v + v_m). \tag{3}$$

The first term is a person's "exogenous" benefit of subscribing. The second term is the subscription price (or rental) charged by operator i . The final term represents the utility a person gets from calling people on-the-go, both from home and when on-the-go. In both cases, the number of such calls is proportional to the measure of consumers also subscribed to the mobile network.

⁵For our calibrated linear demands model, our general results continue to hold even though we account for the possibility demands can become zero.

⁶The most important implication of this additive benefits assumption for our results is that the marginal subscriber receives the same number of calls as the average subscriber. This is consistent with empirical evidence from Australia, where it has been found that new customers receive close to the average amount of *FTM* calls.

We assume mobile operators face a constant per minute cost of terminating *FTM* calls on a mobile network, denoted c , and a cost per subscriber they attract, denoted f . A two-stage framework is adopted. In the first stage, each mobile operator i sets a *FTM* termination charge a_i .⁷ In the second stage, the fixed-line and mobile operators set their prices. There are $m \geq 2$ mobile operators that are homogenous price competitors.

Discriminatory pricing by the fixed-line operator is allowed so that p_f can be different for calling customers on different mobile networks. Assuming the per-minute cost of originating *FTM* calls is C , the perceived cost for calling a customer on operator i 's network is $C + a_i$. For simplicity, it is assumed the *FTM* price to network i is $C + a_i$. The assumption is that retail prices are regulated or competitive resale is allowed over the fixed-line's network with access granted at cost.

Finally, we assume the price of *MTM* calls is set at true cost. One possible explanation is that mobile operators agree to terminate each others calls at costs and given they set two-part tariffs it will then be optimal for them to price *MTM* calls at marginal cost. As a result, in our setting firms will not discriminate prices between on and off-net *MTM* calls.⁸ If, instead, the termination charge for *MTM* calls was set above cost then off-net prices will be more expensive, and given we assume mobile operators are homogenous price competitors, this would mean in equilibrium everyone would subscribe to a single operator. This would again result in the *MTM* price being set equal to cost.

3 Analysis and results

This section characterizes the equilibrium and welfare maximizing termination charges. We first characterize these for the general case, before doing so for a number of special cases.

Competition implies only the operator which sets the lowest possible rentals will obtain any demand.

⁷This assumes mobile operators can make take-it-or-leave-it offers, motivated by the fact the fixed-line operator is typically obligated (by regulation) to interconnect with the mobile operators. Binmore and Harbord (2005) provide an analysis of mobile termination instead as a bilateral monopoly bargaining problem, which will generally result in lower termination charges being set than those characterized here.

⁸Allowing for imperfectly competitive networks, Gans and King (2001) show that in equilibrium firms will set a reciprocal termination charge below cost, so as to soften downstream competition. The result would be lower off-net prices than on-net prices. However, if anything, the reverse pattern of prices is typically observed.

Since all operators are symmetric, this implies

$$r = f - \pi_T, \quad (4)$$

where π_T denotes *FTM* termination profit per-subscriber. There will be a subsidy to mobile customers in the sense their subscription price will be set below subscription costs f . Equivalently, a subsidy on mobile phone purchases may exist, which in practice can be amortized into the subscription price.

For a particular termination charge a , the termination profit per subscriber is

$$\pi_T = (1 - N)\pi_M + N\pi_C, \quad (5)$$

which is the weighted sum of the termination profit per subscriber $\pi_M = (a - c)q_f$ when people calling from home have no choice of technology (the standard assumption) and the termination profit per subscriber $\pi_C = (a - c)\bar{q}_f$ when they do have a choice. The weights are the proportion of consumers that do not subscribe and the proportion that subscribe. Given $q_f > \bar{q}_f$, we have $\pi_M > \pi_C$ for $a > c$ so that $\pi_M > \pi_T$. For the sake of obtaining the general results in Proposition 1 below we assume π_M , π_C and π_T are concave in a over the relevant range of a , a property which we will illustrate is true in our calibrated model of Section 4.

To determine the demand N for mobile subscription, define the critical value of b (denoted b^*) above which all people choose to subscribe to a mobile network and below which they do not. The number of people that will subscribe is then $1 - G(b^*)$. At b^* the utility in (3) just equals that in (2). This implies

$$b^* = r - (v - v_f + v_m)N. \quad (6)$$

The demand for mobile subscription is then determined by solving (6) together with the definition

$$N = 1 - G(b^*). \quad (7)$$

For demand to be well behaved, some restriction on the size of network effects has to be imposed. If these network effects are too strong then demand becomes explosive (a slightly lower rental rate can lead to 100% penetration.) Attracting one additional mobile subscriber increases the value of being able to make *MTM* calls so much that it leads to more than one additional subscriber. These additional subscribers then attract further subscribers, with this process snowballing until all potential subscribers

have joined. Throughout, it will be assumed that network effects are not so strong that they make demand explosive in this way. This assumption is indeed correct for the calibrated model.

The number of mobile subscribers will be an increasing function of the *FTM* termination charge a through two effects. First, a impacts termination profit per subscriber. Higher termination profits per subscriber result in competing mobile operators offering lower subscription prices (the waterbed effect), thereby increasing mobile subscription. Second, a affects the price of *FTM* calls. A higher price of *FTM* calls makes it more attractive to avoid high *FTM* charges by becoming a mobile subscriber, thereby increasing mobile subscription. This second effect has been ignored in the literature to date.

To characterize the equilibrium termination charge, define the following two benchmark termination charges. Denote the termination charge maximizing π_M as a_M , which is characterized by

$$a_M = c + \frac{q_f}{|q'_f|}, \quad (8)$$

where $q'_f = \partial q_f / \partial p_f$. This is the equilibrium termination charge when there is no *FTM* to *MTM* substitution, and the one analyzed in the existing literature.⁹ Secondly, denote the termination charge maximizing π_T with N held constant as a_N , which is characterized by

$$a_N = c + \frac{Gq_f + (1 - G)\bar{q}_f}{G|q'_f| + (1 - G)|\bar{q}'_f|}. \quad (9)$$

This would be the equilibrium termination charge in our framework if when setting their termination charges, mobile operators assume mobile penetration is held fixed, but still allow for *FTM* to *MTM* substitution. Then we have:

Proposition 1 *The equilibrium termination charge a^* is above cost but lower than if mobile operators ignore the impact of their termination charge on mobile penetration, which is in turn lower than the equilibrium termination charge in the absence of *FTM* to *MTM* substitution. That is, $c < a^* < a_N < a_M$.*

Proof. Competition between identical mobile operators implies mobile operators will set their termination charge to maximize π_T . If others do not do so, setting their termination charge to give a higher

⁹Assuming no *FTM* to *MTM* substitution, it is the same termination charge that a mobile operator would set if it was the only operator in the mobile sector (it has a monopoly), or in fact, regardless of the nature of competition in the mobile sector (see Wright, 2002 for an analysis allowing for quite general forms of competition between mobile operators).

level of π_T than their rivals will allow an operator to set a rental below their rivals, thereby capturing the whole market while still covering their costs. In equilibrium, at least one operator must set a_i at this level. Moreover, in equilibrium people will only subscribe to operators who set their termination charge at this level. It therefore follows that the equilibrium termination charge a^* is the one that maximizes π_T .

Termination profits π_T must be maximized for $a > c$ since by increasing a slightly above c ensures a positive level of π_T . This implies $a^* > c$.

Competition between operators implies all termination profit is competed away in the second stage. Substituting (4), (5) and (7) into (6), the marginal consumer is defined implicitly by

$$b^* = f - (a - c) (Gq_f + (1 - G)\bar{q}_f) - (1 - G) (v - v_f + v_m). \quad (10)$$

Totally differentiating (10) gives

$$\frac{db^*}{da} = - \frac{(q_f + (a - c) (Gq'_f + (1 - G)\bar{q}'_f))}{1 - g\eta}, \quad (11)$$

where the network effects are $g\eta$ and

$$\eta = v - v_f + v_m - (\pi_M - \pi_C).$$

Given our assumption that network effects are not explosive, it must therefore be that $g\eta < 1$. Given π_M and π_C are concave in a , the function $G\pi'_M + (1 - G)\pi'_C$ is decreasing in a so that $G\pi'_M + (1 - G)\pi'_C > 0$ for $a < a_N$. This implies that $db^*/da < 0$ for $c \leq a \leq a_N$.

Differentiating π_T with respect to a gives the first order condition

$$\pi'_T = G\pi'_M + (1 - G)\pi'_C + g \frac{db^*}{da} (\pi_M - \pi_C). \quad (12)$$

Since $\pi_M > \pi_C$ (given $q_f > \bar{q}_f$) for $a > c$, and $db^*/da < 0$, then $\pi'_T < G\pi'_M + (1 - G)\pi'_C$. Since π_T is concave, π'_T is decreasing and equal to zero only at a^* . This implies $\pi'_T = 0$ must occur for $a < a_N$. That is, $a^* < a_N$.

Solving implicitly for a^* implies

$$a^* = c + \left(\frac{1}{1 + \phi_E} \right) \frac{Gq_f + (1 - G)\bar{q}_f}{G|q'_f| + (1 - G)|\bar{q}'_f|} \quad (13)$$

where

$$\phi_E = \frac{g(1-G)(\pi_M - \pi_C)(q_f - \bar{q}_f)}{(1-g\eta)(Gq_f + (1-G)\bar{q}_f) - g(\pi_M - \pi_C)q_f}. \quad (14)$$

Since $c < a^* < a_N$, it must be that $\phi_E > 0$.

Finally, to show $a_N < a_M$, we compare the associated first order conditions $\pi'_M = q_f + (a-c)q'_f = 0$ and $G\pi'_M + (1-G)\pi'_C = 0$. Since π_M is concave in a , π'_M is decreasing, and so the result $a_N < a_M$ follows provided $G\pi'_M + (1-G)\pi'_C < \pi'_M$. This is true since $\pi'_C < \pi'_M$ which holds for $a > c$ given (1) and that $q_f > \bar{q}_f$. ■

Proposition (1) shows that the effect of allowing *FTM* to *MTM* substitution relative to existing models where this effect is ignored is to lower the equilibrium *FTM* termination charge and retail price. This happens for two reasons that can be best seen by examining (12). First, assume the number of mobile subscribers remains constant as the termination charge changes so the additional effect of $db^*/da < 0$ is ignored. The resulting equilibrium termination charge is a_N as characterized in (9). In this case *FTM* demand is more elastic compared to the case without substitution, which is why a_N is lower than a_M . Second, noting that the number of mobile subscribers increases as the termination charge increases, the equilibrium termination charge will be even lower than a_N . This is reflected by the additional negative term in (12). As the termination charge is increased, more people will become mobile subscribers to avoid high *FTM* charges, which will further reduce the demand for *FTM* calls, thereby constraining equilibrium termination charges to be lower than a_N .

Now compare this with the termination charge that maximizes consumer surplus and that maximizes welfare. Since in equilibrium, profits of both the fixed-line and mobile operators are zero, welfare, being defined as the sum of consumer and producer surplus, just equals consumer surplus.¹⁰ Given a symmetric second-stage equilibrium in which all mobile operators set the same rental r , welfare can be written

$$W = \int_{b^*}^{\bar{b}} (b - r + N(v + v_m) + (1-N)v_f)g(b)db.$$

Total welfare is the expression in (3) summed up over all mobile subscribers plus the expression in (2) multiplied by the measure of people without a mobile subscription. Using that $r = f - \pi_T$, and $N = 1 - G$,

¹⁰Alternatively, if C is to be interpreted to include the margins of the fixed-line operator then our welfare measure needs to be interpreted as just corresponding to consumer surplus.

we get

$$W = \int_{b^*}^{\bar{b}} (b - f + G\pi_M + (1 - G)\pi_C + (1 - G)(v + v_m) + Gv_f)g(b) db, \quad (15)$$

which is what we seek to maximize. For the sake of obtaining the general results in Proposition 2 below we assume functional forms are such that W is concave in a over the relevant range of a , a property which we will again illustrate is true in our calibrated model of Section 4.

Proposition 2 *The welfare maximizing termination charge a_W is above cost. In general, the equilibrium termination charge a^* may be higher or lower than a_W .*

Proof. Differentiating (15) with respect to a gives the first order condition, which can be written as

$$W' = (1 - G)(\pi'_T - Gq_f - (1 - G)\bar{q}_f) - g\frac{db^*}{da}(Gv_f + (1 - G)(v + v_M)). \quad (16)$$

Ignoring the last term by setting $db^*/da = 0$, we get $W' < 0$ for $a = a^*$ since at this termination charge $\pi'_T = 0$. Given W is assumed to be concave in a , this implies $a_W < a^*$. Incorporating the last term will not change this result when db^*/da is close to zero at a^* (it equals zero in the approach taken in the existing literature), or when the hazard rate is sufficiently low. However, given $db^*/da < 0$ from Proposition 1, the last term is positive and this means it is also possible for $W' > 0$ at $a = a^*$, so that $a_W > a^*$. This case arises when db^*/da is sufficiently negative and the hazard rate is sufficiently high (a numerical example is supplied in Section 4).

Setting (16) equal to zero and using (11), after considerable rearrangement we get

$$a_W = c + \left(\frac{1 + \delta}{1 + \phi_S}\right) \frac{Gq_f + (1 - G)\bar{q}_f}{G|q'_f| + (1 - G)|\bar{q}'_f|}, \quad (17)$$

where

$$\delta = \frac{(1 - G)(q_f - \bar{q}_f)}{Gq_f + (1 - G)\bar{q}_f} > 0$$

and

$$\phi_S = \frac{1 - g\eta}{\rho v_f + g\eta}. \quad (18)$$

Recall $g\eta < 1$ for demand to be well behaved. We also have $\eta > 0$ since $v + \pi_C > v_f + \pi_M$ and $v_m > 0$. The first inequality is true given the general property that welfare is higher when a single (monopoly) firm sells to a consumer which faces the choice to buy a substitute good at cost than a consumer without such a choice. Thus, $\phi_S > 0$ and so $a_W > c$. ■

Proposition 2 shows that the welfare maximizing termination charge depends on an inverse elasticity rule, as does the equilibrium termination charge a^* . Moreover, the welfare maximizing termination charge exceeds cost, a result that contradicts the cost based approach adopted by regulators in a number of countries including the United Kingdom. Setting the termination charge above cost is needed to subsidize mobile subscription, which is socially optimal given additional mobile subscribers generate a positive externality to fixed-line callers and to other mobile subscribers. This result is consistent with findings from earlier studies as is discussed below.

Taking into account *FTM* to *MTM* substitution also implies a key result from the earlier literature, that the equilibrium termination charge is always too high, is no longer true. Earlier works, emphasized that starting from a_M , a small decrease in the termination charge must increase welfare since it has no first-order impact on termination profits and so on the number of mobile subscribers, but leads to a first-order benefit to *FTM* callers by reducing the deadweight loss from *FTM* prices being set above marginal cost. Allowing for *FTM* to *MTM* substitution changes this logic since a small decrease in the termination charge below a^* does lead to a first-order impact on the number of mobile subscribers. As the *FTM* price falls, fewer people will obtain a mobile subscription since they have less need to do so to avoid high *FTM* prices. Since these mobile subscribers generate positive externalities on others, welfare may be lowered.

To compare the equilibrium termination charge with that which maximizes welfare, several special cases can be considered. We start with the benchmark case considered previously in the literature and explain the new effects our approach uncovers.

3.1 Comparisons with existing literature

The existing approaches used to explore *FTM* termination, such as Armstrong (1997, 2002), Gans and King (2000), Thompson *et al.* (2006), Valletti and Houpis (2005) and Wright (1999, 2002), assume there are no *MTM* calls, and therefore no scope for *FTM* to *MTM* substitution. The case without *MTM* calls is just a special case of our framework, obtained by imposing that $\bar{q}_f(p_f, p_m) = q_f(p_f)$ and $\bar{q}_m(p_f, p_m) = q_m(p_m) = 0$. Imposing these conditions on (13) implies $a^* = a_M$. In equilibrium, competing mobile operators set their termination charge at the same level as would be chosen by a monopolist operator. Thus, ignoring *FTM* to *MTM* substitution results in the equilibrium termination charge being overstated as in Proposition 1. Allowing for *FTM* to *MTM* substitution removes the “monopoly outcome” result and demonstrates the importance of competition that exists for the majority of calls to people on-the-go.

Imposing the same conditions on (17) implies the welfare maximizing termination charge is characterized by

$$a_W = c + \left(\frac{1}{1 + \phi_S} \right) \frac{q_f}{|q'_f|}, \quad (19)$$

where

$$\phi_S = \frac{1}{\rho v_f}.$$

This implies a_W is still above cost but now necessarily lower than a^* . Moreover, it also depends on the inverse elasticity rule, although only after appropriate discounting of the monopoly markup.¹¹

The result in (19) is missing several important effects that arise once *MTM* calls are incorporated. Allowing for *MTM* calls, but not substitution with *FTM* calls, the discount to the monopoly markup becomes

$$\phi_S = \frac{1 - gv_m}{\rho v_f + gv_m}.$$

The existence of *MTM* calls leads to positive network effects, represented by the term gv_m , which

¹¹This result is of the same type found in the optimal Ramsey pricing literature for regulated industries where the regulated price is compared to the unregulated monopoly price (Brown and Sibley, 1986, pp. 40-41).

suggests a higher termination charge than (19) is socially optimal.¹² In contrast, in (19) the only reason for setting the termination charge above cost is the positive externality on *FTM* callers. In other words, the existing approach understates the extent to which the welfare maximizing termination charge should be set above cost. More generally, the result implies that the more important are network externalities introduced by $v_m > 0$ and $v_f > 0$, the closer is a_W to a_M . However, the welfare maximizing termination charge remains between c and a_M , as in the existing literature, a result obtained in previous models by assuming away *MTM* calls. The analysis here shows it is not necessary to do so.

If we also allow for *FTM* to *MTM* substitution, we get three additional effects not previously considered. Mobile subscribers will now not make so many *FTM* calls, given they can substitute to *MTM* calls instead, thereby reducing the “tax base” which means a higher “tax rate” or termination charge is socially optimal. This is captured in the positive δ term in the numerator of (17). Second, network effects are now different. The network effects consist of the benefits v_m arising to mobile subscribers from now being able to make calls to more people who are on-the-go (the additional mobile subscribers). If mobile subscribers do not substitute between *FTM* and *MTM* calls, this effect would be the only network effect. However, attracting additional mobile subscribers now leads to two new effects reflected by the two additional terms in η . The first of these effects is that additional mobile subscribers make it more attractive for non-subscribers to join the mobile network to make *MTM* calls to avoid high *FTM* prices since they increase the number of calls that can be made to people on-the-go, which is the second term in η . On the other hand, additional mobile subscribers reduce the tax base of the mobile operators, thereby reducing the subsidy available to mobile subscribers, which reduces the size of the network effect, explaining the third term in η . The second term in η is always larger than the third, as explained in the proof of Proposition 2, so that the network effects are stronger as a result of allowing mobile subscribers to substitute between *FTM* and *MTM* calls. Finally, compared to (19), *FTM* demand is now more elastic, reflecting the ability of mobile subscribers to substitute to *MTM* calls, which means a lower level of termination charge is socially optimal but for the same reason the equilibrium level is also lower.

Another way of comparing our results to those of the existing literature, is to examine whether the

¹²This can be confirmed by comparing first order conditions, with the assumption that the original welfare function is concave in a .

welfare maximizing termination charge is closer or further away from the equilibrium termination charge once *MTM* calls and *FTM* to *MTM* substitution is taken into account. To evaluate this we focus on the ratio of markups

$$\frac{a^* - c}{a_W - c}.$$

Evaluating this before taking into account *MTM* calls, this ratio equals

$$1 + \frac{1}{\rho v_f},$$

which exceeds one. After taking these features into account, the ratio becomes equal to

$$\frac{1 + \phi_S}{(1 + \phi_E)(1 + \delta)},$$

where ϕ_S , defined in (18), will generally be lower than $1/\rho v_f$ and ϕ_E defined in (14) is positive (since $c < a^* < a_N$). Since $\delta > 0$, the ratio will generally be lower than before, suggesting the extent to which the equilibrium charge exceeds the welfare maximizing level is now lower. In fact, as Proposition 2 established, it is possible for the ratio to be lower than one, with the equilibrium termination charge then being too low.

3.2 New markets

Consider the situation in a relatively immature market, where in equilibrium few people subscribe initially. Perhaps this situation reflects that the private benefits of joining are small to start with, so \bar{b} is low. Mobile operators will not attract many subscribers even though doing so generates positive externalities. This scenario can be captured by considering the limit case in which $G = 1$ (so $N = 0$) but with $g > 0$, so that there are no mobile subscribers in equilibrium even though some people would subscribe if their utility could be increased a little bit. Then imposing $G = 1$ on (13) and (17) implies that $a^* = a_W = a_M$. In equilibrium, the termination charge is set to the monopoly level, reflecting that there is no substitution to *MTM* calls given no one has a mobile phone subscription in equilibrium. However, this outcome is also the welfare maximizing termination charge given it maximizes termination profits, thereby providing the greatest incentive to encourage mobile subscription, without which consumers get no surplus from *FTM* calls.

3.3 Mature market

At the other extreme is the case in which everyone subscribes to the mobile network. This might reflect a situation where having a mobile phone is viewed as a necessity, and so even if mobile subscribers are not subsidized, all will continue to subscribe. Provided the penetration rate is fixed so that we have $g = 0$ and $G = 0$, then the equilibrium charge is defined by $a^* = c + \bar{q}_f / |\bar{q}'_f|$, and the welfare maximizing termination charge equals termination cost c . Note a^* can be quite close to cost c , since \bar{q}_f could be highly elastic (if *MTM* calls are quite close substitutes to *FTM* calls). The outcome in this market setting can be quite different if the marginal subscriber remains sensitive to the rentals even as the penetration rate approaches one. We explore this case in terms of the calibrated model in the next section.

4 Calibrated model

In this section, the model is calibrated to Australian data. Linear demands will be used throughout since they afford analytical solutions which makes the model particularly suitable for use in policy settings where a range of different parameter values may want to be considered.¹³

We first fix the underlying parameter values to match observed or plausible values of various variables. The mobile penetration rate N in 2003 in Australia is taken as 0.72 from Australian Communications Authority (Telecommunications Performance Report 2002-03, December 2003. p. 90.) The per-minute price of *FTM* calls p_f is taken as \$0.33, the per-minute price of *MTM* calls p_m is taken as \$0.10, and the corresponding *FTM* termination charge a is taken as \$0.21 (ACCC Mobile Services Review, June 2004 and data from Australian mobile company websites). While p_f and a are directly observed, the price of *MTM* calls is taken as the average price over all minutes, and reflects that many packages involve on-net *MTM* calls being free. This approach could be motivated by an assumption that callers only consider the average price they pay since they do not always know which network the person they call is on. The corresponding average monthly rental for mobile subscription is set at \$22 (data from 2004 for Australian

¹³The calibration below is available from the authors in a MS Excel file that allows users to change variable settings and consider different scenarios. With a non-linear specification, calibration results will generally depend on having appropriate starting values. We are currently attempting to do the calibration exercise with a generalized CES demand specification.

mobile operators).

A major challenge for the calibration exercise is that the four underlying components of demand q_f , \bar{q}_f , \bar{q}_m and q_m are not observed. Instead, we observe the total number of calls received per mobile subscriber, broken down only into *FTM* calls and *MTM* calls. Denoting the total *FTM* demand as Q_f and the total *MTM* demand as Q_m , the relationship between the individual components of demand and observed demand is¹⁴

$$Q_f = (1 - N) q_f + N \bar{q}_f$$

and

$$Q_m = N (q_m + \bar{q}_m).$$

These are measured as the total number of minutes called per mobile subscriber per month. Using data from an Australian carrier we have that $Q_f = 25.687$ minutes and $Q_m = 66.713$ meaning that the typical mobile subscriber receives more than twice as many calls from people making *MTM* calls than *FTM* calls.¹⁵

The price elasticity of *FTM* demand per mobile subscriber is defined as $E_f = - (dQ_f/dp_f) (p_f/Q_f)$. This has been estimated to be as low as 0.6 (ACCC op. cit., p. 103) and as high as 2 (using data from an Australian mobile carrier and econometrics estimates by the authors). We set the elasticity at the mean value of 1.3 but also consider these higher and lower values.¹⁶ We note in passing that the ACCC

¹⁴With the demand specification assumed below, the proportion of time people are at “home” λ is irrelevant to the final results provided it lies strictly between 0 and 1. It can however be useful in interpreting the demand equations. For this reason it is set to 0.85, which is our estimate based on U.S. and Australian survey data for the proportion of waking hours when individuals have access to a fixed line.

¹⁵Interestingly, the pattern of calling is the reverse in the U.S. According to the CTIA SemiAnnual Wireless Survey, June 2005, 70% of the calls received by a typical subscriber are *FTM* calls and only 30% are *MTM* calls. This outcome is exactly what one would expect if there is a lot of substitution to *MTM* calls in Australia to avoid the high price of *FTM* calls but little such substitution in the U.S. where the price of a *FTM* calls is the same as a local call.

¹⁶In taking the derivative dQ_f/dp_f to match this elasticity, we assume N is held constant. This is an approximation since a one percent increase in p_f may induce a small change in N , and this in turn a small change in Q_f to the extent that $q_f > \bar{q}_f$. This approximation is equivalent to the actual elasticity being slightly different from the one matched. For the benchmark case in Table 1, the elasticity matched is actually 1.28% rather than 1.3%.

elasticity, which is similar to elasticities that other regulatory bodies such as UK Ofcom have used in their findings, is inconsistent with the claim that each mobile operator is a monopolist with respect to *FTM* calls. An operator with a monopoly over termination of calls to its subscribers will always set its termination price to the point where *FTM* demand becomes elastic. Unlike earlier models of competitive bottlenecks, our model, which permits substitution between *FTM* and *MTM* calls, can handle inelastic demand and we therefore explore the implications of low price elasticities in the sensitivity analysis below. Further, in the absence of other information, it is assumed that the price elasticity of $q_m(p_m)$ with respect to p_m is equal to the price elasticity of $q_f(p_f)$ with respect to p_f when measured at the same price p_f since in both cases callers have no alternative way to call the person they wish to call, and in both cases they are calling the same type of person — a person on-the-go.

To identify the individual components of demand, some additional assumptions are required. The preference of mobile subscriber at home in terms of making *FTM* versus *MTM* calls in order to call someone on-the-go is specified in terms of the proportion of each type of call they will make and how willing they are to substitute between the two types of calls. For the former, we fix the ratio of *FTM* to *MTM* calls for a mobile subscriber at home when both types of calls are assumed to be free. This ratio, denoted ψ , is set equal to four. The assumption is that if all calls are free and with both technologies available, people will call four times as much using their landline as compared to their mobile phone, perhaps reflecting that fixed-line calls are perceived to be of higher quality or safer. Secondly, we assume that the cross-price elasticity of demand (\bar{q}_f with respect to p_m) is equal to some fraction of the magnitude of the own-price elasticity of demand (\bar{q}_f with respect to p_f), when evaluated at equal prices. This fraction, denoted γ , is set equal to one half.

Finally, in fitting the distribution function we make use of the estimated elasticity of penetration $E_r = -(dN/dr)(r/N)$ which equals 0.55 from Hausman (1997). Similar values have been estimated for Australia (Hausman, 2002).

Tables 1 summarize the levels of different variables and parameters that allow us to calibrate all the parameters of the model. The first ten variables are either observed or based on estimates, while the remaining two are based on plausible guesses. For the estimated elasticities and the last two unknown parameters, a wide range of plausible values is allowed for as shown in the final row of the table.

Variables	N	p_f	p_m	a	r	Q_f	Q_m	λ	E_r	E_f	ψ	γ
Base value	0.72	0.33	0.1	0.21	22	25.7	66.7	0.85	0.55	1.3	4	0.5
Range									± 0.2	± 0.7	± 3	± 0.2

Table1. Values of parameters for calibration.

Note. All monetary units are in Australian dollars.

The parameters c and C can be obtained directly from the above numbers. The model implies $p_m = 2c$ (assuming the same cost of originating and terminating mobile phone calls), so that $c = 0.05$. This is also broadly consistent with estimates of the cost of terminating mobile calls (ACCC op. cit. and NZ Commerce Commission, Investigation into Regulation of Mobile Termination, Draft Report, October 2004). This configuration implies the cost (markup) parameter for *FTM* calls is $C = p_f - a = 0.12$.¹⁷

The linear demand for call minutes corresponds to representative consumers having the following quadratic utility function¹⁸

$$U(\bar{q}_f, \bar{q}_m) = \alpha(\bar{q}_f + \omega\bar{q}_m) - \frac{\beta}{2}(\bar{q}_f^2 + 2\gamma\bar{q}_f\bar{q}_m + \bar{q}_m^2).$$

A mobile subscriber at home maximizes $U(\bar{q}_f, \bar{q}_m) + y$ subject to $p_f\bar{q}_f + p_m\bar{q}_m + y \leq M$, where y is expenditure on all other goods and M is income. This leads to the following demand functions

$$\begin{aligned}\bar{q}_f(p_f, p_m) &= \frac{\alpha(1 - \omega\gamma) - p_f + \gamma p_m}{\beta(1 - \gamma^2)} \\ \bar{q}_m(p_f, p_m) &= \frac{\alpha(\omega - \gamma) - p_m + \gamma p_f}{\beta(1 - \gamma^2)},\end{aligned}$$

provided p_f and p_m are sufficiently close, where using our earlier definitions $\bar{q}_f = \lambda\bar{q}_f(p_f, p_m)$ and $\bar{q}_m = \lambda\bar{q}_m(p_f, p_m)$.

If instead $p_f \leq (p_m - \alpha(\omega - \gamma))/\gamma$ then $\bar{q}_f(p_f, p_m) = q_f(p_f) = (\alpha - p_f)/\beta$ and $\bar{q}_m(p_f, p_m) = 0$ which is the demand for *FTM* calls of non-subscribers. The same functional form is adopted for *MTM*

¹⁷Based on regulatory proceedings in the US and Australia, the cost of *FTM* origination is thought to be around 3 cents per minute. This suggests the fixed line operator has a margin of around 9 cents per-minute. For now, we treat $C = 0.12$ as the true cost of origination thereby ignoring this margin, to remain consistent with our theoretical model. Another interpretation is that this is the fixed-line operator's profit margin, and so by ignoring it, our welfare measure really just measures consumer surplus. This latter possibility is explored further below.

¹⁸In a slight abuse of notation, we temporarily use \bar{q}_f and \bar{q}_m in the utility function to denote the consumer's choice variables even though they are not multiplied by λ as defined earlier.

calls for people on-the-go, who also have no substitution possibilities, so that $q_m(p_m) = (\alpha_m - p_m) / \beta_m$. Alternatively, if $p_f \geq \alpha(1 - \omega\gamma) + \gamma p_m$ then $\bar{q}_f(p_f, p_m) = 0$ and $\bar{q}_m(p_f, p_m) = (\alpha\omega - p_m) / \beta$ which happens when the termination charge is set high enough. The corresponding indirect utility measures $v = \lambda v(p_f, p_m)$, $v_f = \lambda v_f(p_f)$ and $v_m = (1 - \lambda) v_m(p_m)$ are obtained in the usual way.

Using that $\psi = \bar{q}_f(0, 0) / \bar{q}_m(0, 0) = 4$ implies $\omega = 2/3$. With ω fixed, the parameter α is determined by matching the elasticity $E_f = 1.3$ so that

$$\alpha = \frac{(1 + E_f)(1 - \gamma^2(1 - N))p_f - \gamma p_m E_f N}{E_f(1 - \gamma^2(1 - N) - \omega\gamma N)} = 0.7348.$$

With α and ω fixed, the parameter β is determined by the condition $Q_f = 25.7$, so that

$$\beta = \lambda \frac{(\alpha - p_f)(1 - \gamma^2(1 - N)) - \gamma(\omega\alpha - p_m)N}{(1 - \gamma^2)Q_f} = 0.0104.$$

The property that the price elasticity of $q_m(p_m)$ with respect to p_m is equal to the price elasticity of $q_f(p_f)$ with respect to p_f when measured at the same price p_f implies $\alpha_m = \alpha$. The parameter β_m is then pinned down by the fact $Q_m = 66.7$ so that

$$\beta_m = \frac{\beta(1 - \lambda)(1 - \gamma^2)(\alpha_m - p_m)N}{Q_m\beta(1 - \gamma^2) + (p_m - \gamma p_f - \alpha(\omega - \gamma))\lambda N} = 0.0013.$$

With these parameter values we find $q_f = 33.0466$, $\bar{q}_f = 22.8430$, $\bar{q}_m = 20.4072$, $q_m = 72.2317$, $v_f = 6.6878$, $v = 8.6005$, $v_m = 22.9245$, $\pi_M = 5.2875$ and $\pi_C = 3.6549$ so that $\pi_T = N\pi_C + (1 - N)\pi_M = 4.1120$. The subsidy per mobile subscriber at these termination charges is slightly more than \$4 per month. (Note π_T can be rewritten as $(a - c)Q_f$, and so only depends on the assumptions on a , c and Q_f and not on the other parameter values.) From this and the observed r , we can back out f which is

$$f = r + \pi_T = 26.1120.$$

Finally, we need to characterize the distribution function for b . We adopt the uniform distribution on $[\underline{b}, \bar{b}]$, since it is consistent with a linear demand for mobile participation in the rental r , and like the rest of the model allows for an analytical solution.¹⁹ The distribution function is

$$G(b) = \frac{b - \underline{b}}{\bar{b} - \underline{b}}.$$

¹⁹We have considered other distributions such as Weibull and log-normal but without truncation they imply the expected value of b is unreasonably high. With truncation, the fitted distributions reasonable something close to a uniform distribution but are then subject to numerical problems.

Using this distribution, we can solve (7) and (6) to get

$$N = \frac{\bar{b} - r}{\bar{b} - \underline{b} - (v - v_f + v_m)} = 0.72$$

so that

$$E_r = \frac{r}{\bar{b} - r} = 0.55.$$

These two equations are solved for \underline{b} and \bar{b} implying

$$\underline{b} = r - (v - v_f + v_m) - \frac{(1 - N) r}{N E_r} = -18.3928$$

and

$$\bar{b} = \left(\frac{1 + E_r}{E_r} \right) r = 62.$$

A negative value of \underline{b} is necessary to explain why 28 percent of consumers still do not have a mobile subscription despite mobile subscription providing greater caller benefits than the charges consumers face. It implies that some people's aversion to having a mobile subscription is greater than their net benefits from using a mobile phone. Still, in the calibrated solution, the marginal subscriber who is just willing to subscribe still has a positive value of b^* , and so the fact some non-subscribers have negative values of b does not affect the equilibrium or welfare maximizing termination charge. Finally, network effects, measured by $g\eta$, equal 0.2886 which implies non-explosive but positive network effects as assumed in the theoretical analysis.

With all parameters determined, we then work out the equilibrium rental, penetration rate, termination profit and welfare for different values of the termination charge a , by solving (7) and (10) for N to get

$$N = \frac{\bar{b} - f + (a - c) q_f}{\bar{b} - \underline{b} + (a - c) (q_f - \bar{q}_f) - (v - v_f + v_m)},$$

where q_f , \bar{q}_f , v and v_f are all functions of a through the *FTM* price $P = C + a$. Figure 1 plots termination profits in the absence of substitution π_M and with substitution π_T , the penetration rate N and total welfare W , as a function of the *FTM* termination charge a . This illustrates the relationship between cost-based termination, the equilibrium termination charge, the monopoly termination charge, the welfare maximizing termination charge, and the termination charge which maximizes mobile penetration.

— Figure 1 to appear here —

As expected from Proposition 1, the introduction of *FTM* to *MTM* substitution causes the equilibrium termination charge to fall below the monopoly one predicted by the traditional analysis. The equilibrium level of a predicted by the calibrated model (that maximizes π_T) is 25 cents, somewhat higher than the 21 cents level observed in the data in June 2004. This seems reasonable since there was likely some regulatory pressure on operators in Australia to lower their charges by 2004.²⁰ In comparison, the level of a predicted by existing competitive bottleneck theories is 33 cents, which is considerably higher than the equilibrium level.

Starting from the equilibrium level of 25 cents, an increase in termination charges increase mobile penetration even though the subsidy per mobile subscriber falls, since by raising the *FTM* price it makes mobile subscription attractive as a way to avoid high *FTM* prices. We find mobile penetration is maximized by a termination charge set at 30 cents, higher than the level which maximizes termination profit, consistent with the idea that high *FTM* prices also encourage mobile participation. Despite this finding, for this benchmark case, we find welfare maximized at the lower level of 18 cents for our calibrated model, reflecting that we have chosen *FTM* demand to be relatively price elastic and mobile subscription to be relatively inelastic with respect to rentals paid.²¹ While lower than the equilibrium level, this level is still more than three times the assumed cost of 5 cents of terminating calls. Furthermore, regulating termination charges to cost results in lower welfare than the unregulated equilibrium. Thus, cost-based regulation of termination charges would be undesirable in this setting. Finally, note that with our calibrated parameter values, one would reach the opposite conclusion if the monopoly termination charge predicted by existing theory were instead used as the equilibrium charge.

²⁰Regulatory hearings into *FTM* termination charges in Australia began in 1999. By 2001 carriers agreed to gradually decrease their termination prices.

²¹For instance, if the price elasticity of *FTM* demand E_f is reduced to 0.6 as used by the ACCC, then the welfare maximizing termination charge increases to \$0.32. If instead, the elasticity of subscription demand E_r is increased to 1, then the welfare maximizing termination charge is \$0.21, exactly the observed level. If E_r is increased above 2.3 then $a_W > a^*$, consistent with the prediction of Proposition 2 that $a_W > a^*$ is possible.

4.1 Sensitivity analysis

A sensitivity analysis is conducted to see how much the calibrated results change when the underlying assumptions are changed. The range of values considered is determined by that given in Table 1. The results are presented in Table 2. The table shows the equilibrium, monopoly and welfare maximizing termination charges if the calibration had instead been conducted to match the alternative values given in the first column.

	a^*	a_M	a_W
Benchmark	0.25	0.33	0.18
$E_r = 0.35$	0.26	0.33	0.16
$E_r = 0.75$	0.25	0.33	0.20
$E_f = 0.6$	0.39	0.53	0.32
$E_f = 2$	0.21	0.27	0.14
$\psi = 1$	0.24	0.41	0.23
$\psi = 7$	0.26	0.32	0.18
$\gamma = 0.3$	0.26	0.28	0.16
$\gamma = 0.7$	0.25	0.46	0.22

Table 2. Equilibrium, monopoly and welfare maximizing termination charges.

Benchmark values of parameters are $E_r = 0.55$, $E_f = 1.3$, $\psi = 4$ and $\gamma = 0.5$.

A number of interesting observations arise from Table 2. The first thing to note is the qualitative findings are quite robust to different assumptions about the values of the different identifying variables. The biggest variation arises from changes to the assumption about the price elasticity of *FTM* demand. If it is instead set at the regulator's estimate of 0.6, the monopoly termination charge becomes very high (53 cents). With the price elasticity of *FTM* demand being low when measured at a termination charge of 21 cents, mobile operators will raise termination charges to the point where demand becomes elastic. This implies both a^* and a_M are high. However, by the same token, it also implies the welfare maximizing termination charge is high, in fact higher than the observed level at the time.

Note once we allow for two or more parameter values to be changed at the same time, one can easily

find cases where the welfare maximizing termination charge exceeds the equilibrium level. For instance, taking all the different combinations of two of the parameters in table 2 being at their extreme values, this is true if $E_r = 0.75$ and $\psi = 1$, if $E_r = 0.75$ and $\gamma = 0.7$, if $E_f = 0.6$ and $\psi = 1$, if $E_f = 0.6$ and $\gamma = 0.7$, and if $\psi = 1$ and $\gamma = 0.7$. Thus, the possibility that the equilibrium termination charge may be too low cannot be ruled out — in fact, it can arise for plausible parameter values.

4.2 Comparisons with previous literature

To compare the results here to those that would be obtained using earlier models, suppose as is standard in existing analysis, *MTM* calls are ignored altogether. Starting with the benchmark parameter values above, and imposing that $\gamma = \psi = p_m = v_m = 0$, the equivalent model with no *MTM* calls is obtained. As expected this implies $a^* = a_M = 0.33$ since the competitive bottleneck is not checked by *FTM* to *MTM* substitution. On the other hand, $a_W = 0.14$, so that the equilibrium termination charge is still more than twice the welfare maximizing level. However, compared to the equilibrium outcome, welfare is higher if termination charges are regulated at cost. This highlights the role *MTM* calls and *FTM* to *MTM* substitution can play in narrowing the gap between privately set and socially optimal termination charges.

4.3 Strong substitution

At the opposite extreme to the case in which substitution with *MTM* calls is not possible, is the case where such substitution is assumed to be very strong. By strong substitution we mean that over the relevant range of termination charges, mobile subscribers always use their mobile phone to make calls to those on-the-go given that it is cheaper to do so. This situation is likely to become increasingly relevant with the convergence of different technologies.

From a welfare perspective, with high substitutability between types of calls, setting a very high termination charge can become desirable since by raising the price of *FTM* calls, this drives almost everyone to subscribe. This generates positive externalities at relatively little cost since most people can then substitute to *MTM* calls with little loss in utility. For instance, if the degree of substitutability is increased to $\gamma = 0.75$ and the model re-calibrated, such an outcome arises with a_W equal to 83 cents. A

very high termination charge becomes socially optimal, at which point mobile subscribers never use their landline for calls from home. Corresponding to this, mobile penetration is maximized at a termination charge equal to 92 cents, with a near 90% penetration rate. In contrast, the equilibrium termination charge, which is limited by the ability of subscribers to easily substitute to *MTM* calls, remains unchanged at 25 cents. In equilibrium, there is still only limited (rather than strong) substitution between *FTM* and *MTM* calls by subscribers.

4.4 Emerging and mature markets

To study different market stages, from emerging markets to mature markets, we exogenously shift the distribution of b by Δ , so that it lies between $\underline{b} + \Delta$ and $\bar{b} + \Delta$. All other parameters are left the same. The parameter Δ is chosen to generate a range of penetration rates. Table 3 shows the corresponding equilibrium and welfare maximizing termination charges.²² The last column shows the percentage change in welfare from regulating termination charges to cost compared to leaving them at the equilibrium level.

	N	a^*	a_W	% ΔW from $a = c$
$\Delta = -40$	0.021	0.33	0.28	-100%
$\Delta = -30$	0.195	0.31	0.23	-32.86%
$\Delta = -20$	0.370	0.29	0.20	-11.44%
$\Delta = -10$	0.545	0.27	0.19	-6.25%
$\Delta = 0$	0.720	0.25	0.18	-5.01%
$\Delta = 10$	0.895	0.24	0.17	-3.04%
$\Delta = 15$	0.982	0.23	0.17	-2.78%
$\Delta = 20$	1.000	0.23	0.08	2.30%
$\Delta = 30$	1.000	0.23	0.05	2.91%

Table 3. Different degrees of market penetration

Benchmark case is $\Delta = 0$. Penetration rate N measured when $a = 0.21$.

In an immature mobile market, where few people wish to become mobile subscribers to start with,

²²The monopoly termination charge is 33 cents regardless of the value of Δ .

there is little scope for *FTM* to *MTM* substitution. Thus, the equilibrium termination charge will approximately correspond to the monopoly one, as illustrated in the case $\Delta = -40$ in table 3. However, this is also close to being welfare maximizing since with few mobile subscribers, the positive externality from subsidising subscription are more important than any deadweight loss from *FTM* prices being above cost. For instance, when the penetration rate is just 2.1% (at $a = 0.21$), lowering the termination charge to cost (in an attempt to eliminate the deadweight loss of above-cost pricing) will result in no one subscribing to the mobile network, and therefore a complete loss of surplus.

On the other hand, as the penetration rate becomes high for exogenous reason, say because all people would choose to subscribe to a mobile operator at current prices (that is, at $a = 0.21$), then there is no scope for attracting additional subscribers and so no reason to maintain termination charges so high from a welfare perspective. However, this does not necessarily imply cost-based termination charges are socially optimal. As the $\Delta = 20$ case illustrates in Table 3, it could be that the socially optimal termination charge is still above cost since by lowering termination charges to cost, it is no longer the case that everyone will subscribe. This could provide a rationale for leaving termination charges above cost even if there is a 100% penetration in an unregulated equilibrium. Moreover, while the calibrated model predicts the socially optimal termination charge is close to cost in a mature market, it also implies the welfare loss (in percentage terms) of not getting it right is relatively small since everyone can substitute to *MTM* calls.

Another possibility is that marginal mobile subscribers generate fewer incoming calls and so the network benefits of subsidizing their participation diminish as a mobile market becomes mature. This view arises, for instance, if people are heterogenous in terms of how much they call out. Those that remain to join the network have not yet done so because they expect to call out little, and calling levels are highly correlated with calls received. Such a view is inconsistent with our model. In our model, everyone expects to receive the same number of calls, and people are heterogenous only in terms of some fixed benefits of belonging to the network. Moreover, such a view is also inconsistent with empirical evidence from Australia, where it has been found that new customers received pretty much the average amount of *FTM* calls.

4.5 Fixed-line operator makes a profit

Based on regulatory proceedings in the U.S. and Australia, the cost of *FTM* origination is around 3 cents per minute. Since the observed *FTM* price is 33 cents, and the observed *FTM* termination charge is 21 cents, this still leaves 9 cents unaccounted for. The approach above is to attribute this to the cost of the fixed-line operator. Equivalently, this can be interpreted as the fixed-line operator's profit margin, as long as welfare above is interpreted as consumer surplus. Adopting the latter interpretation, the above results imply consumer surplus is maximized at a termination charge of 18 cents per-minute. Attributing a constant 9 cent per-minute margin to the fixed-line operator on *FTM* calls, total welfare is then maximized at 15 cents per-minute. The effect is to lower the welfare maximizing termination charge since starting from the consumer-surplus maximizing termination charge, a lower termination charge increases the total number of *FTM* calls and thereby the fixed-line operator's profit.

This possibility also suggests an alternative regulatory approach to lower *FTM* prices. Here we contrast two policy approaches — one which lowers the fixed-line operator's margin on *FTM* calls (perhaps through allowing resale or direct price regulation), and another which lowers the *FTM* termination charge by the same amount. In contrast to lowering the *FTM* termination charge, lowering the margin on *FTM* calls stimulates the demand for *FTM* calls which increases rather than decreases the tax base for mobile operators. It therefore provides an effective way to reduce the deadweight loss on *FTM* callers without limiting the subsidy to mobile subscribers that provide a positive externality to other callers. For instance, starting from the observed level of the termination charge ($a = 0.21$), a 9 cent drop in a results in a slight increase in total welfare (including the fixed-line operator's profit) of less than 1%.²³ In contrast, a 9 cent cut in the fixed-line operator's margin results in welfare increases by 6.4%.

4.6 Other simplifying assumptions

Two important simplifying assumption of our framework are that mobile operators are homogenous price competitors and that call receivers do not value receiving calls. The first assumption is made for analytical convenience given we allow the number of mobile subscribers to be endogenous. If mobile operators are

²³However, consumer surplus is lower from this change.

instead modeled as offering differentiated services, this may lead to higher welfare maximizing termination charges as a subsidy to the mobile sector helps offset the negative effects of any market power. Whether optimal termination charges are indeed higher also depends on the nature of the pass-through of a subsidy to mobile operators to their customers. With homogenous price competition the rate of pass-through is one. With imperfect competition the pass-through can be more or less than one, depending on the nature of demand. This means results will become more sensitive to the nature of demand.

The second assumption could be relaxed in our framework by adding some utility to mobile subscribers from receiving calls. If callers care about receiving calls, then in a traditional setting this will tend to lower the equilibrium and welfare maximizing termination charge (see Armstrong, 2002 and Wright, 2002). Lower termination charges are set to encourage more *FTM* calls, thereby increasing the utility of mobile subscribers. However, with the possibility of *FTM* to *MTM* substitution, the impact of receivers caring about incoming calls may not be so significant. A high *FTM* termination charge will cause more people to get mobile subscription, which has two offsetting effects on the number of calls received by mobile subscribers. First, it means more people on-the-go that can make calls to a mobile subscriber, thereby increasing the number of calls received by a mobile subscriber. Second, it means more callers that can substitute to cheaper *MTM* calls, so that the higher price of *FTM* calls does not lead to such a big decrease in calls received by mobile subscribers.

To illustrate the impact of these effects, note that in our calibrated model an increase in the termination charge from cost to the equilibrium level reduces calls received from a non-subscriber by 35%. This is the traditional effect emphasized in the literature. However, taking account of mobile subscribers and *MTM* calls, we find the total minutes of calls received by a mobile subscriber only decreases by less than 5%. Thus, even if subscribers are allowed to get some positive utility from receiving calls, this is unlikely to overturn our qualitative findings. Moreover, if people get utility from receiving calls, then the utility fixed-line callers get from receiving calls from people on-the-go should also be taken into account. Such utility can only be realized if people have mobile phones, thereby providing another reason why including the utility of call receivers, is unlikely to overturn our results.

5 Conclusions

A common view among regulators is that fixed-to-mobile prices are set “too high.” This inference follows typically because mobile termination is seen to be a monopoly problem, where mobile carriers do not face competition to terminate calls to their customers. This view misses two important economic factors.

First, some or all of the excess of price over cost for mobile termination will be used to lower prices for mobile subscription, as would be expected in a two-sided market. This will result in more people obtaining a mobile subscription. Further, fixed line callers’ consumer surplus increases when they are able to reach more people “on the go” via their mobile phones, as does that of existing mobile subscribers. Otherwise these people would be unreachable. In other words, every new mobile subscriber attracted acts as a “new good” to the other consumers.

The second economic factor, which this paper considers for the first time, is that consumers who want to reach someone on the go can substitute mobile-to-mobile calls for fixed-to-mobile calls, and will do so given relative prices. With mobile penetration at levels of 70% or higher in most OECD countries, this substitution can take place for most callers. For Australia, we find that mobile subscribers receive over two times as many mobile-to-mobile calls as fixed-to-mobile calls indicating a significant amount of substitution — exactly the reverse pattern is found in the U.S. where there is little, if any, price differential between the two types of calls.

When we allow for substitution to mobile-to-mobile calls, the monopoly termination implication no longer holds true. Indeed, fixed-to-mobile price elasticities used by regulators are inconsistent with the monopoly termination model they have based their regulatory decisions on. Empirically, when we calibrate our model for Australia we find that the monopoly termination rate is much higher than the observed market determined termination rate, so it is very unlikely that the monopoly fixed-to-mobile model, ignoring mobile-to-mobile substitution, is correct. Our model opens up the possibility that the market may in fact set a termination charge that is too low. More importantly, we find that the market determined mobile termination rate is considerably closer to the welfare maximizing termination rate than a regulatory determined termination rate based on cost. Finally, we note that when the mobile penetration is close to 100%, then although in theory welfare can be higher with cost based termination,

it can only be so by a small amount since then everyone can take advantage of mobile-to-mobile calls.

In terms of a policy recommendation we conclude that regulating termination charges on the basis of costs is unjustified. When there is effective competition in mobile markets, and given economic analysis indicates that consumer surplus is higher with market outcomes than cost-based regulation, such market intervention seems misguided. While our calibrated results suggest welfare may be enhanced by lowering termination charges below the market determined rate, they also suggest it is well beyond usual regulatory calculations to determine the optimal prices. Rather, to a first order approximation, regulation of mobile termination rates is doing income redistribution among consumers of relatively small amounts of money. Most economists would agree that income redistribution is not a proper goal of telecommunications regulatory policy.

6 References

Armstrong, M. (1997) “Mobile telephony in the UK,” Regulation Initiative Discussion Paper No. 15, London Business School.

Armstrong, M. (2002) “The theory of access pricing and interconnection,” Chapter 8 in M. Cave, S. Majumdar, and I. Vogelsang (eds.) *Handbook of Telecommunications Economics*, Volume 1, North-Holland.

Armstrong, M. (2006) “Competition in two-sided markets,” *Rand Journal of Economics* forthcoming.

Armstrong, M. and J. Wright (2006) “Two-sided markets, competitive bottlenecks and exclusive contracts,” *Economic Theory* forthcoming.

Binmore, K. and D. Harbord (2005) “Bargaining over fixed-to-mobile termination rates: countervailing buyer power as a constraint on monopoly power,” *Journal of Competition Law and Economics* 1: 49-72.

Brown, S. and D. Sibley. (1986) “The Theory of Utility Pricing,” Cambridge: Cambridge UP.

Caillaud, B. and B. Jullien (2001) “Competing Cybermediaries,” *European Economic Review* 45: 797-808.

Caillaud, B. and B. Jullien (2003) “Chicken and egg: competition among intermediation service providers,” *Rand Journal of Economics* 34: 309-328.

Gans J.S. and S.P. King (2000) "Mobile network competition, customer ignorance and fixed to mobile call prices," *Information Economics and Policy* 12: 301-328.

Gans J.S. and S.P. King and J. Wright (2005) "Wireless Communications," in S.K. Majumdar, I. Vogelsang and M. Cave (eds.), *Handbook of Telecommunications Economics*, Volume 2. Elsevier North-Holland: Amsterdam.

Hausman, J. (1997) "Valuation and the Effect of Regulation on New Services in Telecommunications," *Brookings Papers on Economic Activity: Microeconomics*.

Hausman, J. (2002) "Mobile Telephone," in M. Cave, S.K. Majumdar and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, North Holland.

Littlechild, S. (2006) "Mobile Termination Charges: Calling Party Pays versus Receiving Party Pays," *Telecommunications Policy* forthcoming.

Rochet, J.-C. and J. Tirole. (2003) "Platform competition in two-sided markets," *Journal of the European Economic Association* 1: 990-1029.

Rochet, J.-C. and J. Tirole. (2006) "Defining two-sided markets," *Rand Journal of Economics* forthcoming.

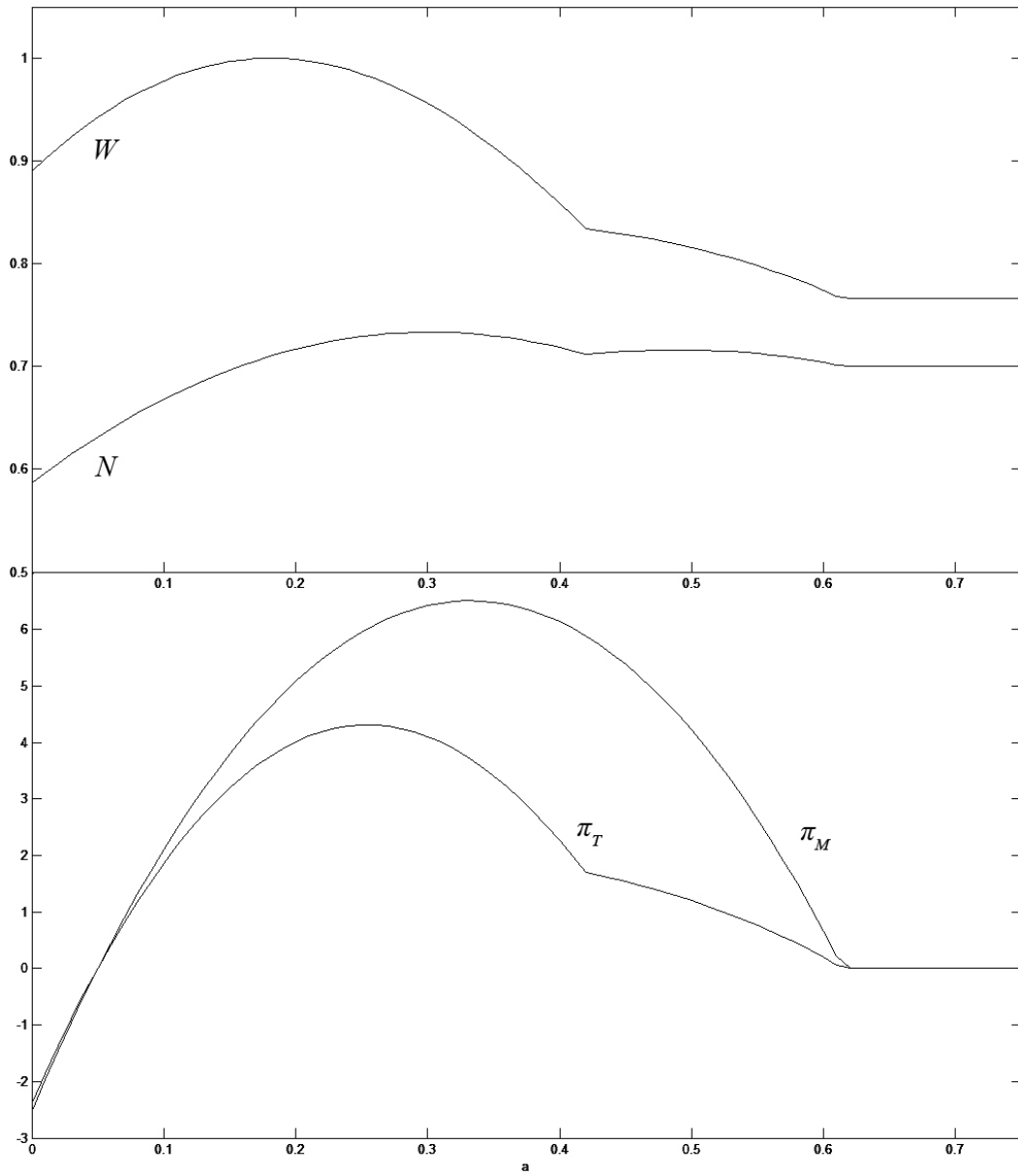
Thomson, H., O. Renard and J. Wright (2006) "Mobile Termination" in R. Dewenter and J. Haucap (eds.) *Access Pricing: Theory, Practice, Empirical Evidence*. Elsevier, forthcoming.

Valletti, T.M and G. Houpis (2005) "Mobile Termination: What is the 'Right' Charge?" *Journal of Regulatory Economics*, 28: 235-258.

Wright, J. (1999) "Competition and Termination in Cellular Networks," Working Paper, Centre for Research in Network Economics and Communications, University of Auckland.

Wright, J. (2002) "Access pricing under competition: An application to Cellular Networks", *Journal of Industrial Economics* 50: 289-315.

Figure 1: Welfare, Penetration Rate and Termination Profits



W = Welfare (normalized)
 N = Penetration rate
 π_M = Termination profits in the absence of substitution
 π_T = Termination profits with substitution